

WHAT IS CLAIMED IS:

1. A method of performing a query in a computer to retrieve data from a database, comprising:

- 5 (a) determining whether a query includes a self join that is transitively derived through table expressions having UNION operators; and
- (b) simplifying the query to eliminate the table expressions and to reduce the query to an equivalent query over tables, when the query includes the self join that is transitively derived through the table expressions having the UNION operators.

10 2. The method of claim 1, wherein the determining step (a) and simplifying step (b) are performed for each quantifier in the query, in order to determine whether the quantifier is part of a self-join that can be eliminated.

15 3. The method of claim 1, wherein the simplifying step (b) comprises: rewriting the table expressions such that the quantified columns to be eliminated are substituted by other quantified columns in a same equivalence class; and once all quantified columns have been rewritten, removing the quantifier from the query.

20 4. The method of claim 3, wherein the equivalence class is determined by the presence of join predicates establishing equality among quantified columns.

25 5. The method of claim 3, wherein the quantifier is only considered for elimination when it is represented in each equivalence class.

6. The method of claim 1, wherein the determining step (a) and simplifying step (b) are implemented as a rule-based transformation, the determining step (a) comprises a condition part of the rule, and the simplifying step (b) comprises an action

part of the rule that is performed when the condition part is true.

7. The method of claim 6, wherein the condition part of the rule requires:

- (1) that a quantifier  $Q_i$  in the query is a FOR EACH quantifier;
- 5 (2) that  $Q_i$  is defined within a SELECT operator that implements an inner join;
- (3) that there are at least two quantifiers  $Q_i$  and  $Q_j$  in the SELECT operator;
- (4) computing equivalence classes among columns in the SELECT operator,

wherein the equivalence classes are divided into three classes:

10 (A) Type 1 equivalence classes are sets of quantified columns over quantifiers other than  $Q_i$ ;

(B) Type 2 equivalence classes are sets of quantified columns that reference  $Q_i$  and at least one other quantifier  $Q_j$ ;

(C) Type 3 equivalence classes are sets of quantified columns that only reference  $Q_i$ ;

15 (5) for each set  $S_i$  of Type 2 equivalence classes, separating quantified columns according to the quantifiers over which they range;

(6) enumerating all combinations of quantified columns  $KEY_1, \dots, KEY_n$  over  $Q_i$ ;

(7) retaining only those  $Q_i$  where  $KEY_i$  comprises key columns;

20 (8) computing the derivation of Superset columns from which  $KEY_i$  is derived and adding them to  $KEY_i$ ;

(9) for each quantifier  $Q_j$  in the SELECT operator where  $Q_j \supset Q_i$ , if  $Q_j$  has a quantified column in each Type 2 equivalence class:

25 (A) computing the derivation of Subset columns from which  $Q_i$  is derived;

(B) for each derivation in  $Q_i$ :

(i) for all lists of quantified columns in the derivation:

(I) the list of quantified columns matches at least one list

of quantified columns in KEY<sub>i</sub>.

8. The method of claim 6, wherein the action part of the rule requires:

(1) for each Type 2 equivalence class:

- 5           (a) replacing all occurrences of quantified columns over Q<sub>i</sub> with  
quantified columns in the same Type 2 equivalence class;  
(2) removing Q<sub>i</sub> from the query.

9. A computer-implemented apparatus for performing a query, comprising:  
10 a computer, wherein the query is performed in the computer to retrieve data from  
a database;

logic, performed by the computer system, for:

- (a) determining whether a query includes a self join that is transitively  
derived through table expressions having the UNION operators; and  
15           (b) simplifying the query to eliminate the table expressions and to reduce  
the query to an equivalent query over tables, when the query includes the self  
join that is transitively derived through the table expressions having the UNION  
operators.

20           10. The apparatus of claim 9, wherein the logic for determining (a) and logic  
for simplifying (b) are performed for each quantifier in the query, in order to determine  
whether the quantifier is part of a self-join that can be eliminated.

25           11. The apparatus of claim 9, wherein the logic for simplifying (b) comprises  
logic for:

rewriting the table expressions such that the quantified columns to be eliminated  
are substituted by other quantified columns in a same equivalence class; and  
once all quantified columns have been rewritten, removing the quantifier from

the query.

12. The apparatus of claim 11, wherein the equivalence class is determined by the presence of join predicates establishing equality among quantified columns.

5

13. The apparatus of claim 11, wherein the quantifier is only considered for elimination when it is represented in each equivalence class.

14. The apparatus of claim 9, wherein the logic for determining (a) and logic for simplifying (b) are implemented as a rule-based transformation, the logic for determining (a) comprises a condition part of the rule, and the logic for simplifying (b) comprises an action part of the rule that is performed when the condition part is true.

15. The apparatus of claim 14, wherein the condition part of the rule requires:

- (1) that a quantifier  $Q_i$  in the query is a FOR EACH quantifier;
- (2) that  $Q_i$  is defined within a SELECT operator that implements an inner join;
- (3) that there are at least two quantifiers  $Q_i$  and  $Q_j$  in the SELECT operator;
- (4) computing equivalence classes among columns in the SELECT operator,

20 wherein the equivalence classes are divided into three classes:

(A) Type 1 equivalence classes are sets of quantified columns over quantifiers other than  $Q_i$ ;

(B) Type 2 equivalence classes are sets of quantified columns that reference  $Q_i$  and at least one other quantifier  $Q_j$ ;

25 (C) Type 3 equivalence classes are sets of quantified columns that only reference  $Q_i$ ;

(5) for each set  $S_i$  of Type 2 equivalence classes, separating quantified columns according to the quantifiers over which they range;

(6) enumerating all combinations of quantified columns KEY1, ..., KEYn over Qi;

(7) retaining only those Qi where KEYi comprises key columns;

(8) computing the derivation of Superset columns from which KEYi is derived  
5 and adding them to KEYi;

(9) for each quantifier Qj in the SELECT operator where  $Qj \supset Q_i$ , if Qj has a quantified column in each Type 2 equivalence class:

(A) computing the derivation of Subset columns from which Qi is derived;

10 (B) for each derivation in Qi:

(i) for all lists of quantified columns in the derivation:

(I) the list of quantified columns matches at least one list of quantified columns in KEYi.

15 16. The apparatus of claim 14, wherein the action part of the rule requires:

(1) for each Type 2 equivalence class:

(a) replacing all occurrences of quantified columns over Qi with quantified columns in the same Type 2 equivalence class;

(2) removing Qi from the query.

20

17. An article of manufacture embodying logic for performing a query in a computer to retrieve data from a database, the logic comprising:

(a) determining whether a query includes a self join that is transitively derived through table expressions having the UNION operators; and

25 (b) simplifying the query to eliminate the table expressions and to reduce the query to an equivalent query over tables, when the query includes the self join that is transitively derived through the table expressions having the UNION operators.

18. The article of manufacture of claim 17, wherein the determining step (a) and simplifying step (b) are performed for each quantifier in the query, in order to determine whether the quantifier is part of a self-join that can be eliminated.

5

19. The article of manufacture of claim 17, wherein the simplifying step (b) comprises:

rewriting the table expressions such that the quantified columns to be eliminated are substituted by other quantified columns in a same equivalence class; and

10 once all quantified columns have been rewritten, removing the quantifier from the query.

20. The article of manufacture of claim 19, wherein the equivalence class is determined by the presence of join predicates establishing equality among quantified  
15 columns.

21. The article of manufacture of claim 19, wherein the quantifier is only considered for elimination when it is represented in each equivalence class.

20 22. The article of manufacture of claim 17, wherein the determining step (a) and simplifying step (b) are implemented as a rule-based transformation, the determining step (a) comprises a condition part of the rule, and the simplifying step (b) comprises an action part of the rule that is performed when the condition part is true.

25 23. The article of manufacture of claim 22, wherein the condition part of the rule requires:

- (1) that a quantifier  $Q_i$  in the query is a FOR EACH quantifier;
- (2) that  $Q_i$  is defined within a SELECT operator that implements an inner join;

(3) that there are at least two quantifiers  $Q_i$  and  $Q_j$  in the SELECT operator;  
(4) computing equivalence classes among columns in the SELECT operator,  
wherein the equivalence classes are divided into three classes:

- 5 (A) Type 1 equivalence classes are sets of quantified columns over  
quantifiers other than  $Q_i$ ;  
(B) Type 2 equivalence classes are sets of quantified columns that  
reference  $Q_i$  and at least one other quantifier  $Q_j$ ;  
(C) Type 3 equivalence classes are sets of quantified columns that only  
reference  $Q_i$ ;  
10 (5) for each set  $S_i$  of Type 2 equivalence classes, separating quantified columns  
according to the quantifiers over which they range;  
(6) enumerating all combinations of quantified columns  $KEY_1, \dots, KEY_n$  over  
 $Q_i$ ;  
(7) retaining only those  $Q_i$  where  $KEY_i$  comprises key columns;  
15 (8) computing the derivation of Superset columns from which  $KEY_i$  is derived  
and adding them to  $KEY_i$ ;  
(9) for each quantifier  $Q_j$  in the SELECT operator where  $Q_j \supset Q_i$ , if  $Q_j$  has a  
quantified column in each Type 2 equivalence class:  
(A) computing the derivation of Subset columns from which  $Q_i$  is  
20 derived;  
(B) for each derivation in  $Q_i$ :  
(i) for all lists of quantified columns in the derivation:  
(I) the list of quantified columns matches at least one list  
of quantified columns in  $KEY_i$ .

25

24. The article of manufacture of claim 22, wherein the action part of the  
rule requires:

- (1) for each Type 2 equivalence class:

- (a) replacing all occurrences of quantified columns over  $Q_i$  with quantified columns in the same Type 2 equivalence class;
- (2) removing  $Q_i$  from the query.

TOP SECRET